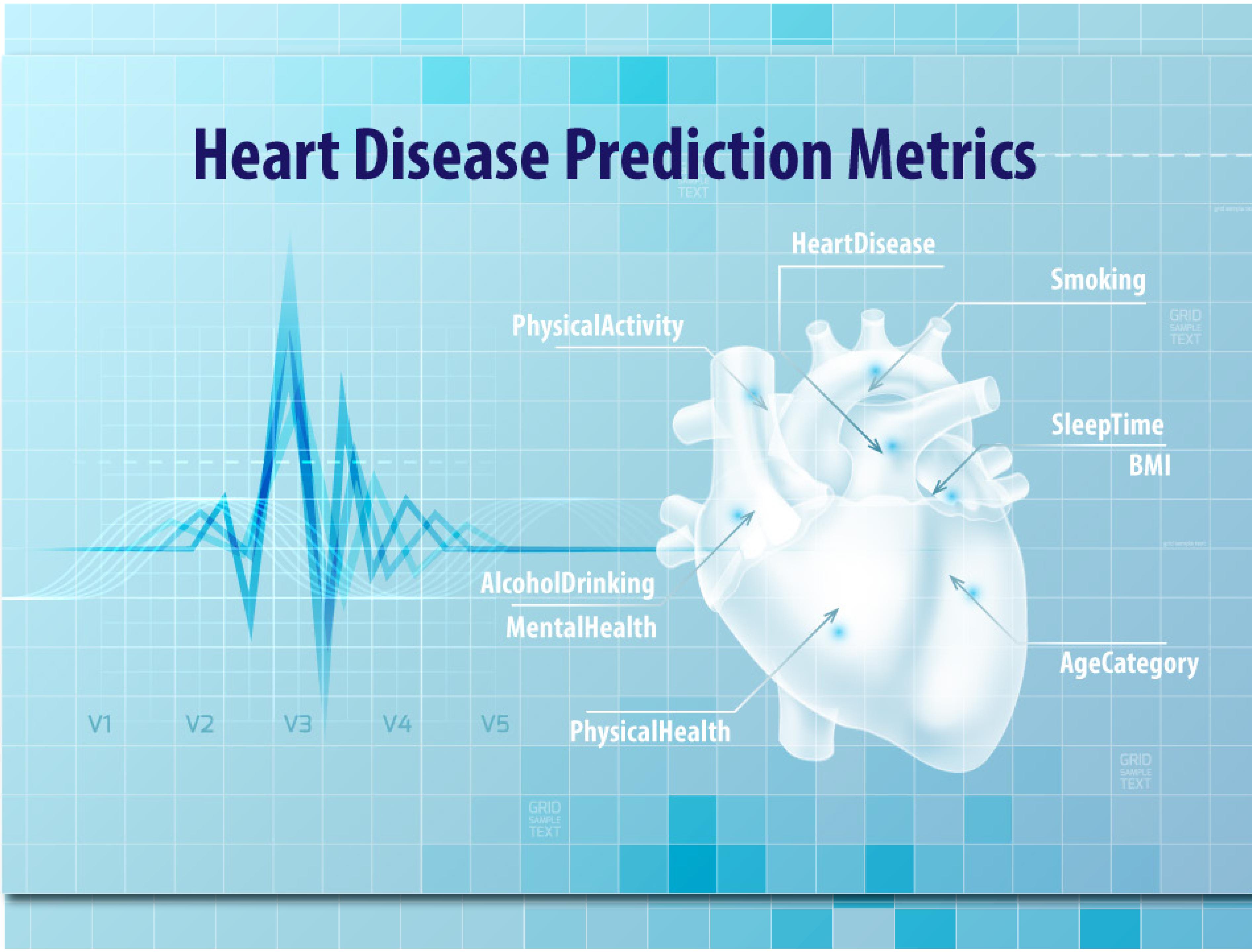


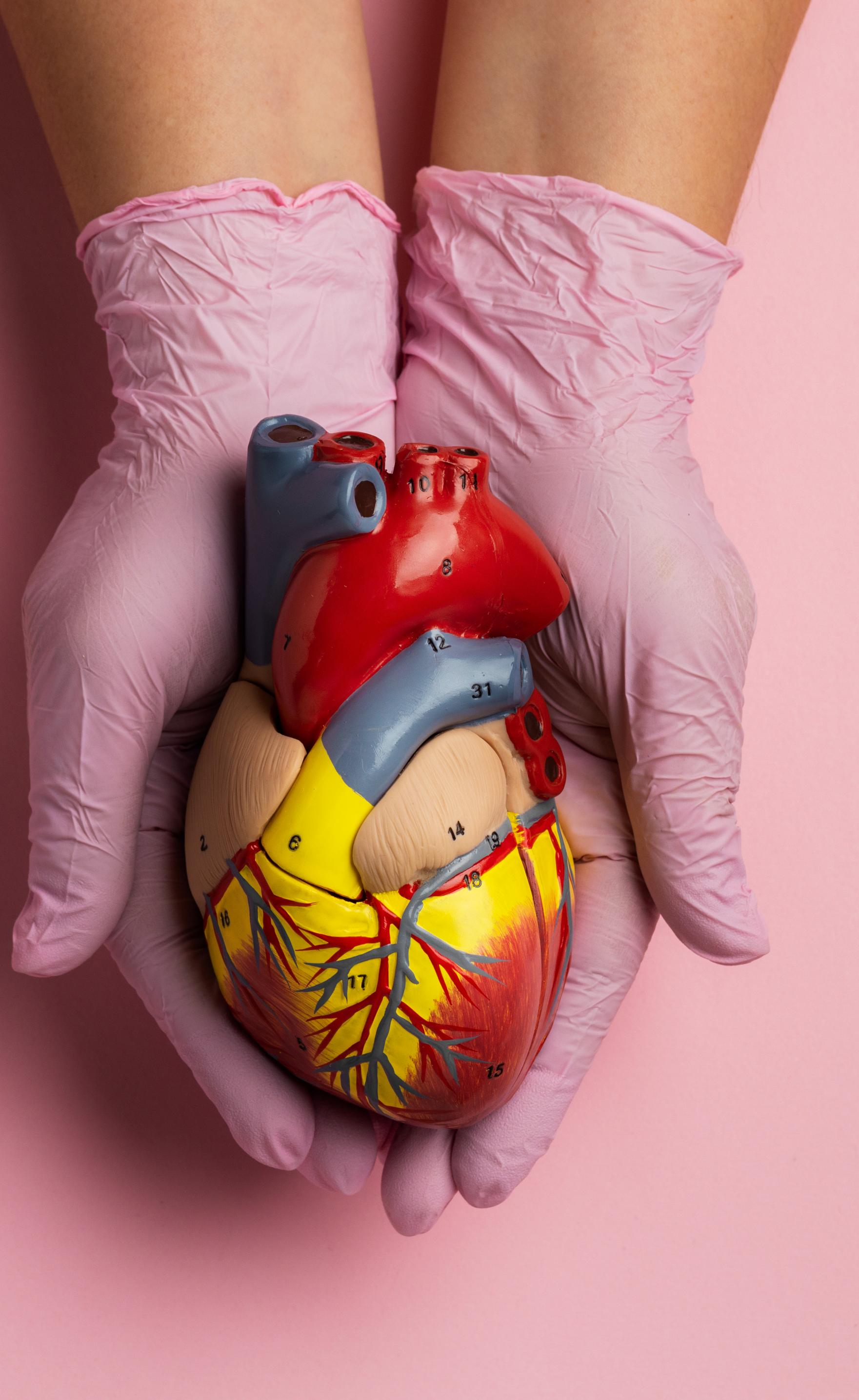
# Heart Disease Prediction Metrics



All Code created by:

- Jesse Kranyak
- Jeff Boczkaja
- Mohamed Altoobi
- Seriesha Mandava

Jan 2024



# Project Purpose Description

This project focuses on the factors that predict heart disease in adults. To determine the factors that predict heart disease in adults we used a single large data set from the [CDC](#). To simplify our workload we utilized a previously cleaned dataset in this project.

The original dataset from the CDC had over 300 data points, and nearly half a million participants spanning decades. Our trimmed dataset from this larger group of data contained 18 data points and 319,795 participants.



# Questions to be addressed

- 1 What is the most important factor in determining heart disease?
- 2 What are the highest correlated factors for heart disease?
- 3 What groups are the most at risk for heart attacks?
- 4 What could be the most important choices made to reverse the potential negative outcomes of heart disease?
- 5 How will this data help hospitals to make informed decisions and develop effective business plans and strategies?

# Overview Of Data Collection

## Cleanup And Exploration Process

**Choose**

a dataset that fit our needs.

1

Explored and understood  
the dataset challenges

2

Used various methods  
to overcome challenges

3

Answered key project  
**questions**

4



```
1 #reframing the data for mapping
2 age_map = {
3     '18-24': 1,
4     '25-29': 2,
5     '30-34': 3,
6     '35-39': 4,
7     '40-44': 5,
8     '45-49': 6,
9     '50-54': 7,
10    '55-59': 8,
11    '60-64': 9
```

Created a comprehensive presentation

5

Aimed to maximize the application of our course learnings throughout the project.



```
1 #loop through all of our columns and see what data they reveal
2
3 def describe_df(df: pd.DataFrame):
4     """
5         This function provides an overview of the dataset including
6             the number of columns, rows, data types, and summary statistics or unique values for each column.
7
8     Parameters:
9         - df: Pandas DataFrame to be described
10
```

10

race\_map = {

'White': 1,  
'Black': 2,  
'Hispanic': 3,  
'Asian': 4,  
'American Indian/Alaskan Native': 5,  
'Other': 6  
}  
  
df['Race'] = df['Race'].map(race\_map)

sex\_map = {  
'Male': 1,  
'Female': 2,  
}

df['Sex'] = df['Sex'].map(sex\_map)

diabetes\_map = {  
'Yes': 3,  
'Yes (during pregnancy)': 2,  
'No, borderline diabetes': 1,  
'No': 0  
}

df['Diabetic'] = df['Diabetic'].map(diabetes\_map)

df = df.drop(['PhysicalHealth'], axis=1)

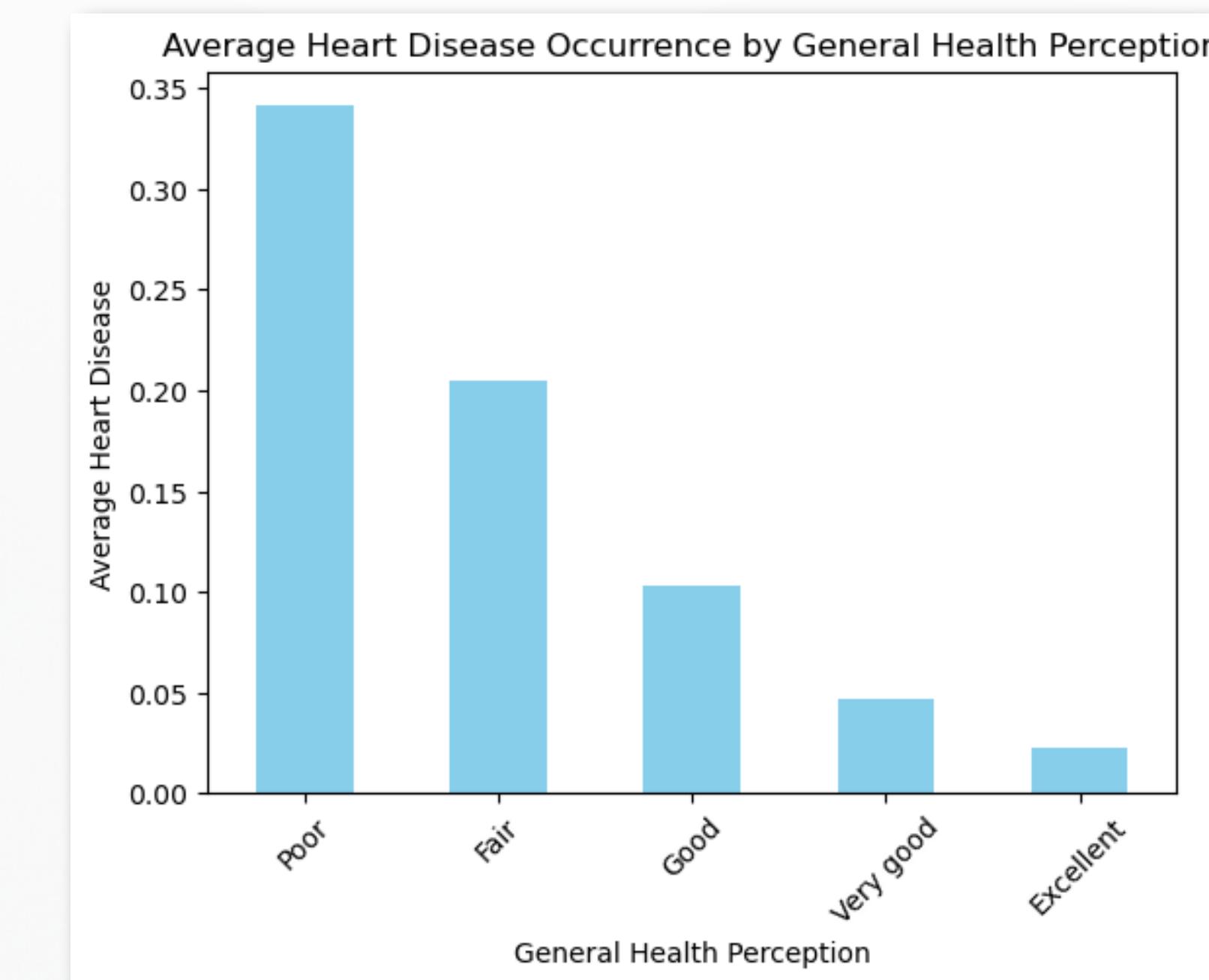
# Q1

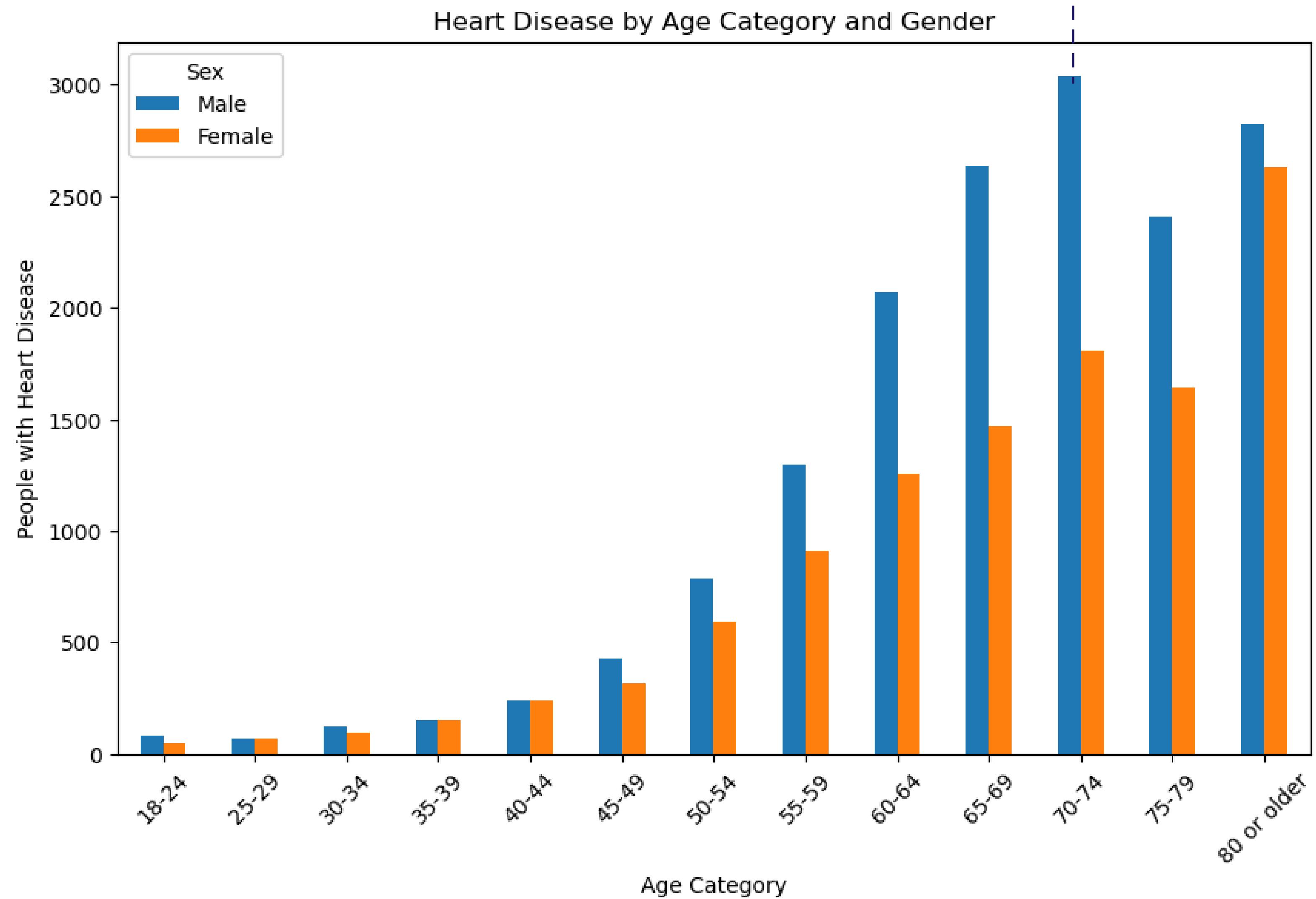
What is the **most important factor** in  
determining heart disease?



For this, we determined that general health was the largest determinant in developing heart disease. You can see from the following graph that even through our lifespan it remains as one of the highest variables correlated to disease. Following general health we can see that physical activity also plays a lifelong role in your heart's health.

Some interesting points we noticed looking at this data are that as we age heart disease becomes more prevalent in men, although men remain at a higher risk than women according to this data, it increases with age.



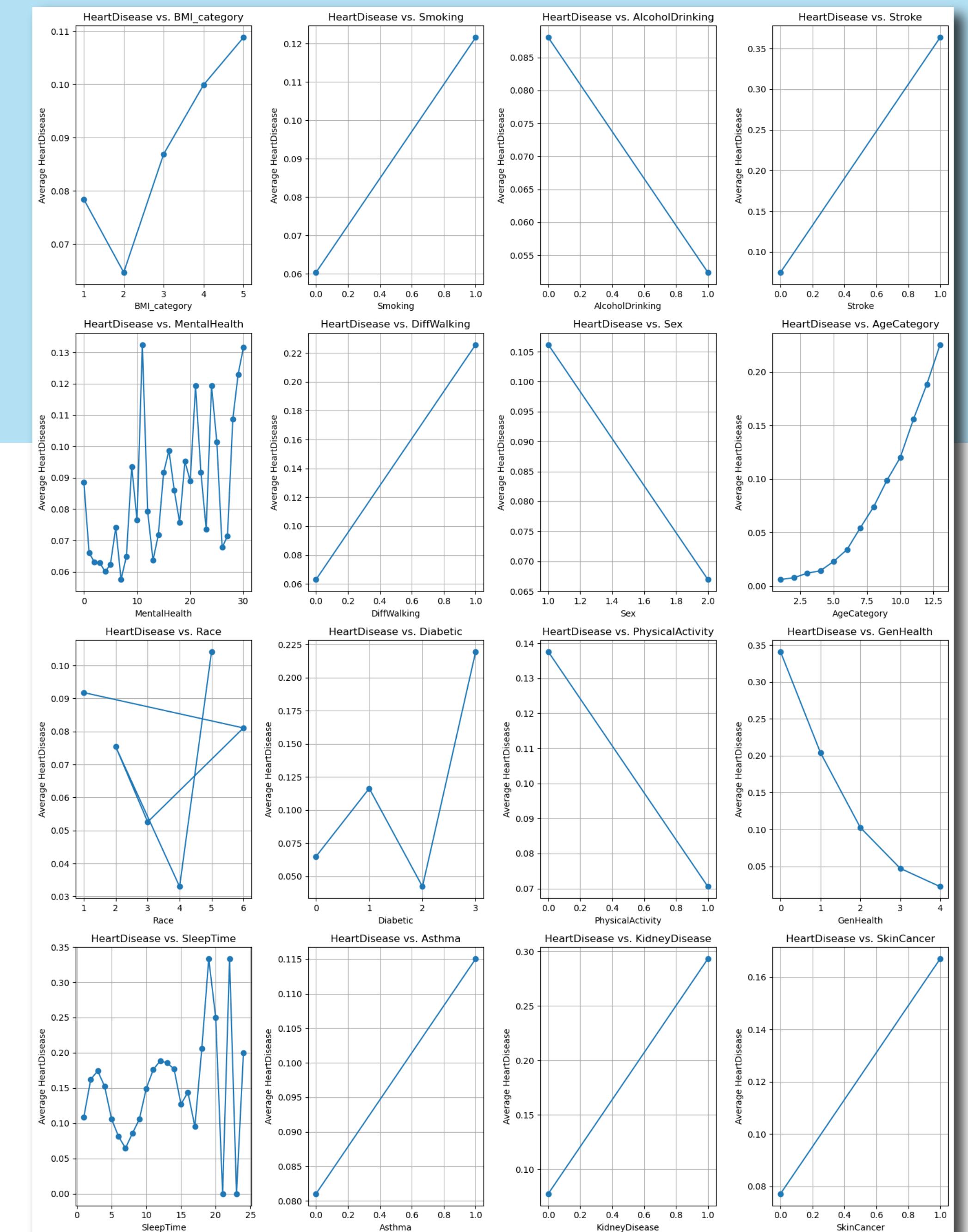
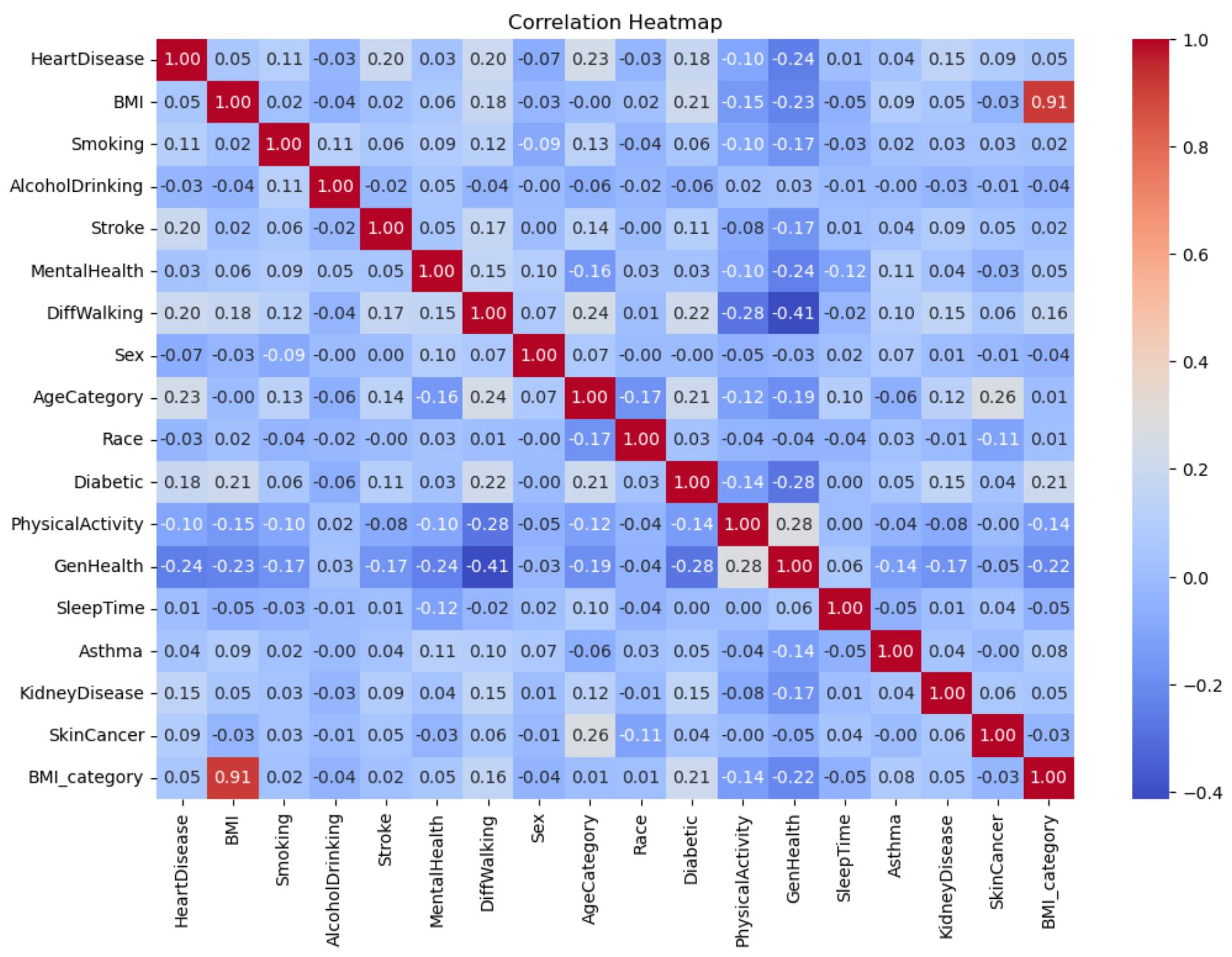




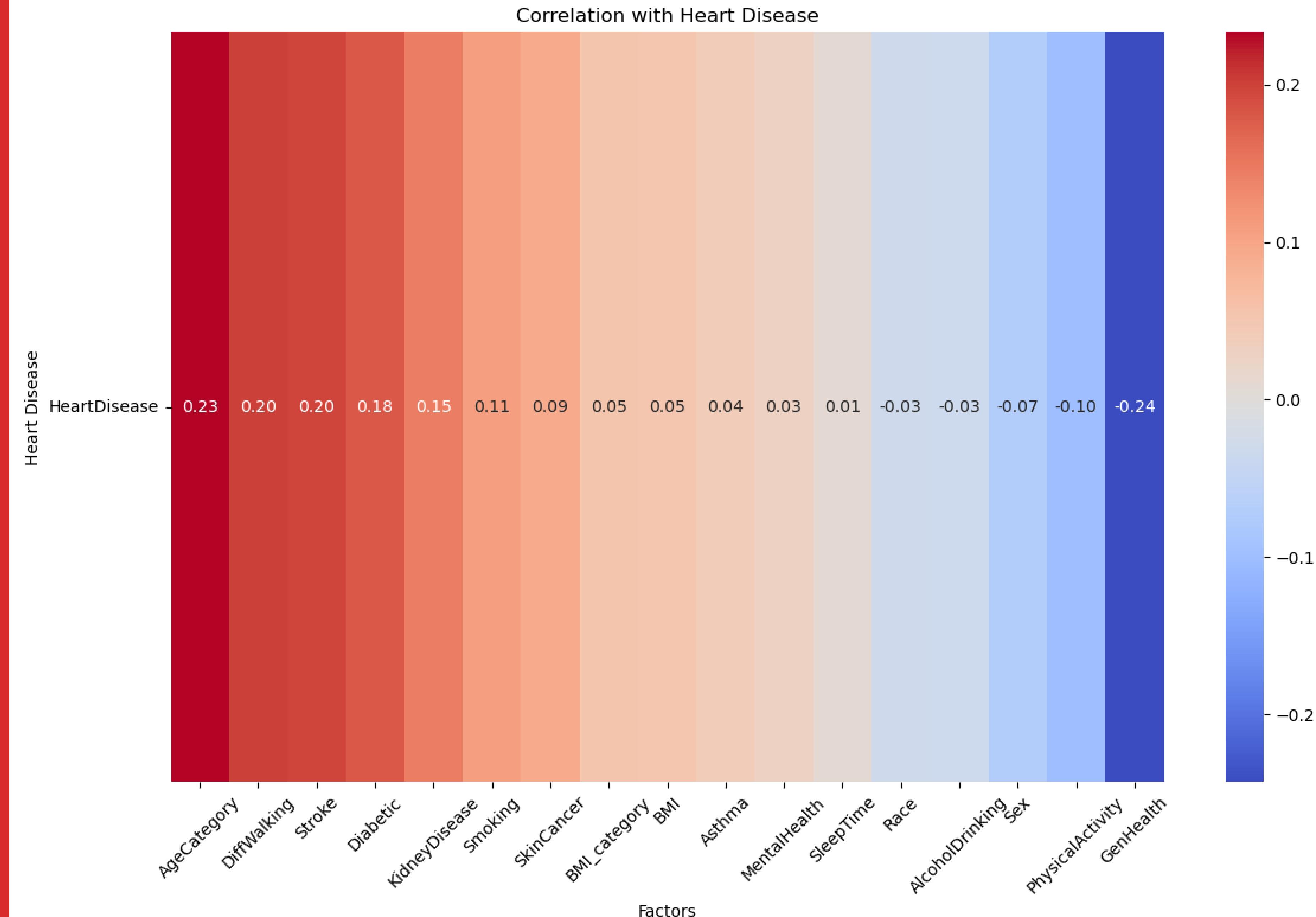
# Q2

What are the **highest correlated factors**  
for heart disease?

We reframed this question to ask; what are the highest correlated factors TO heart disease and ran our data. The highest correlated factors were age and general health, this seemed intuitive. Other than that, some more data we uncovered seemed to also be rather intuitive; such as the correlation between general health and difficulty walking, which had the highest heat map correlation.



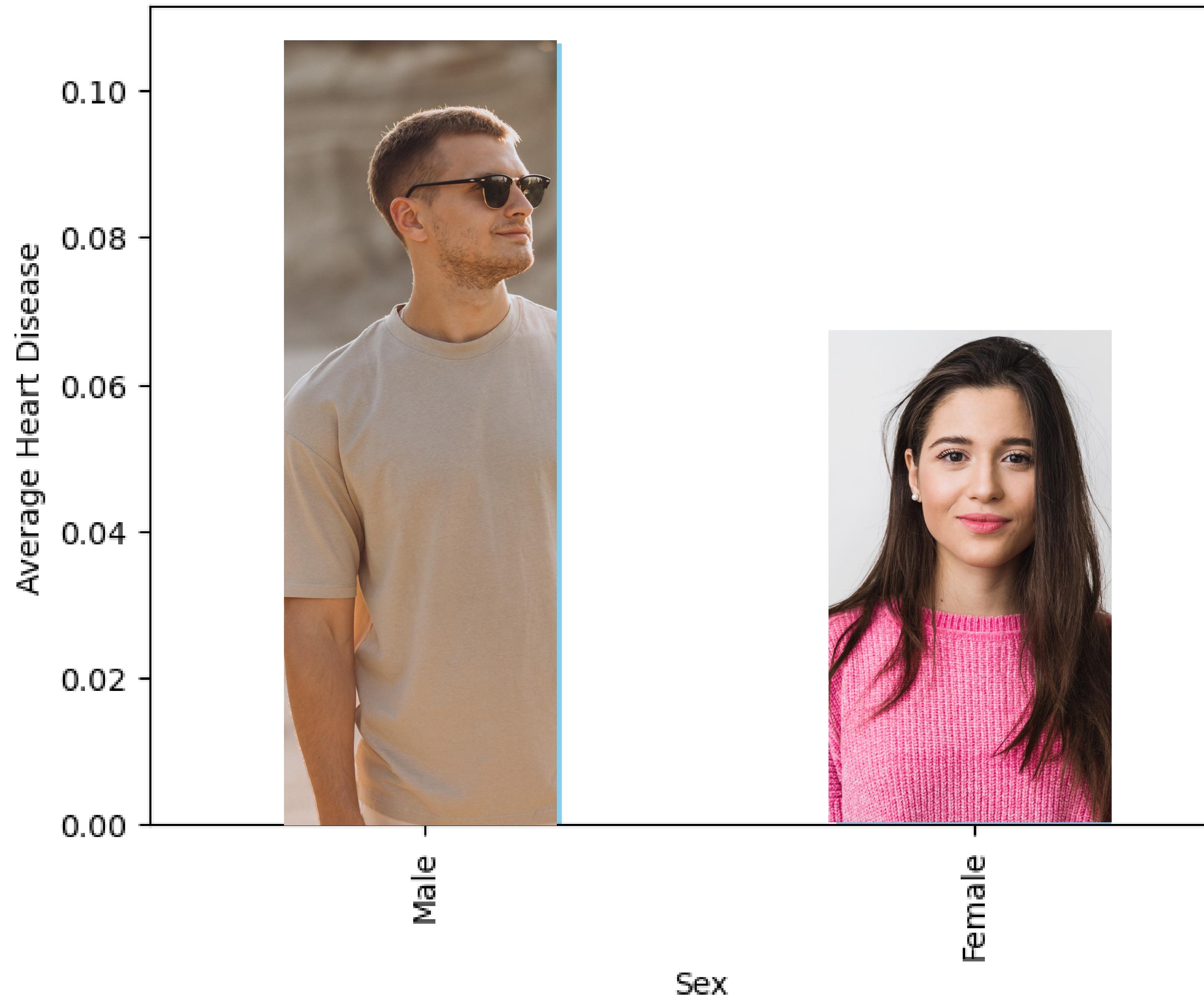
# Correlations And Risk Factors



# Q3

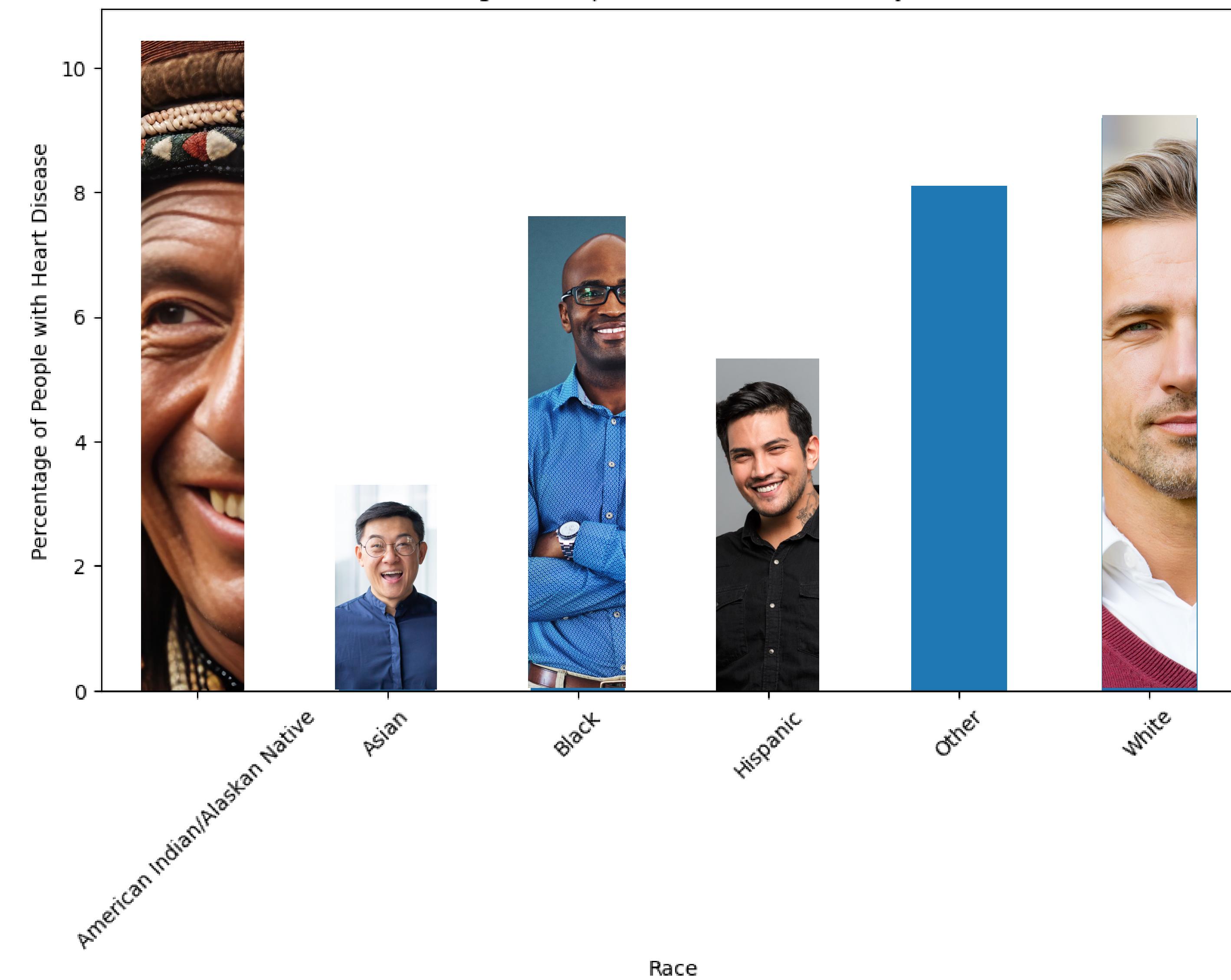
**What groups are the most at risk for heart attacks?**

## Average Heart Disease Occurrence by Sex

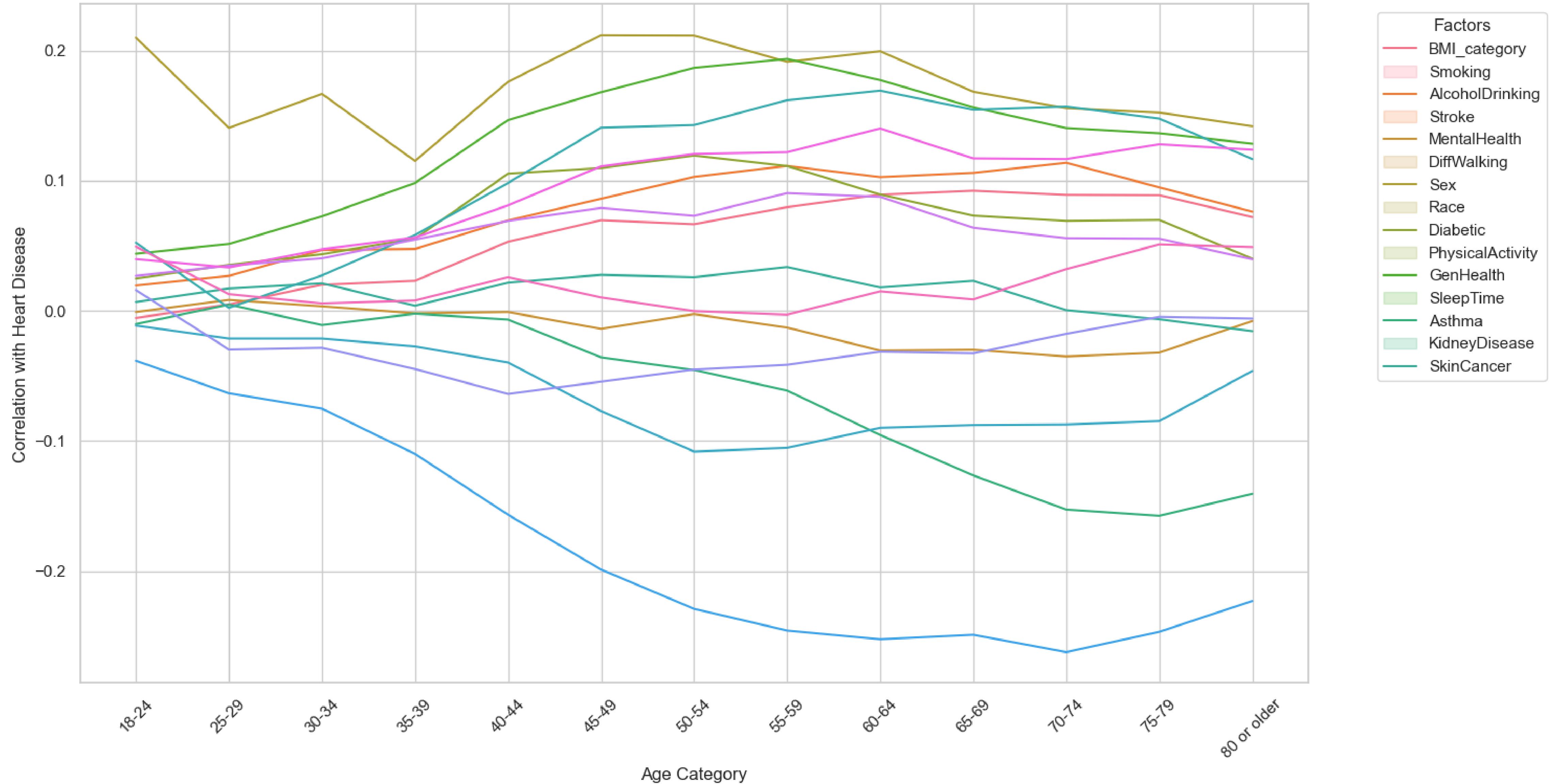


We found that the most at-risk groups for heart disease were aging populations, men, Native Americans followed by Whites, smokers vs non-smokers (12% vs 6%), and having diabetes.

Percentage of People with Heart Disease by Race



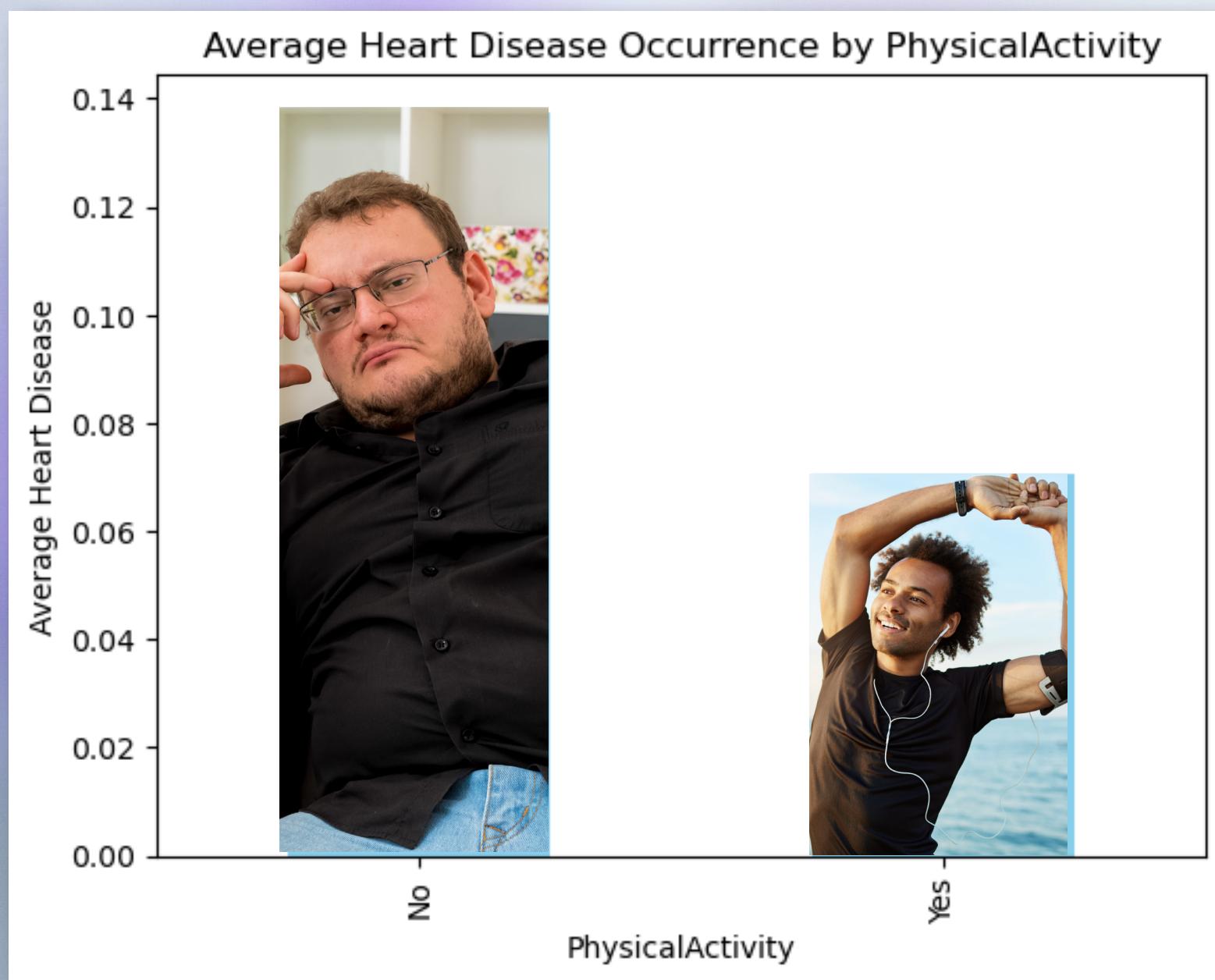
Correlation of Various Factors with Heart Disease Across Different Age Categories

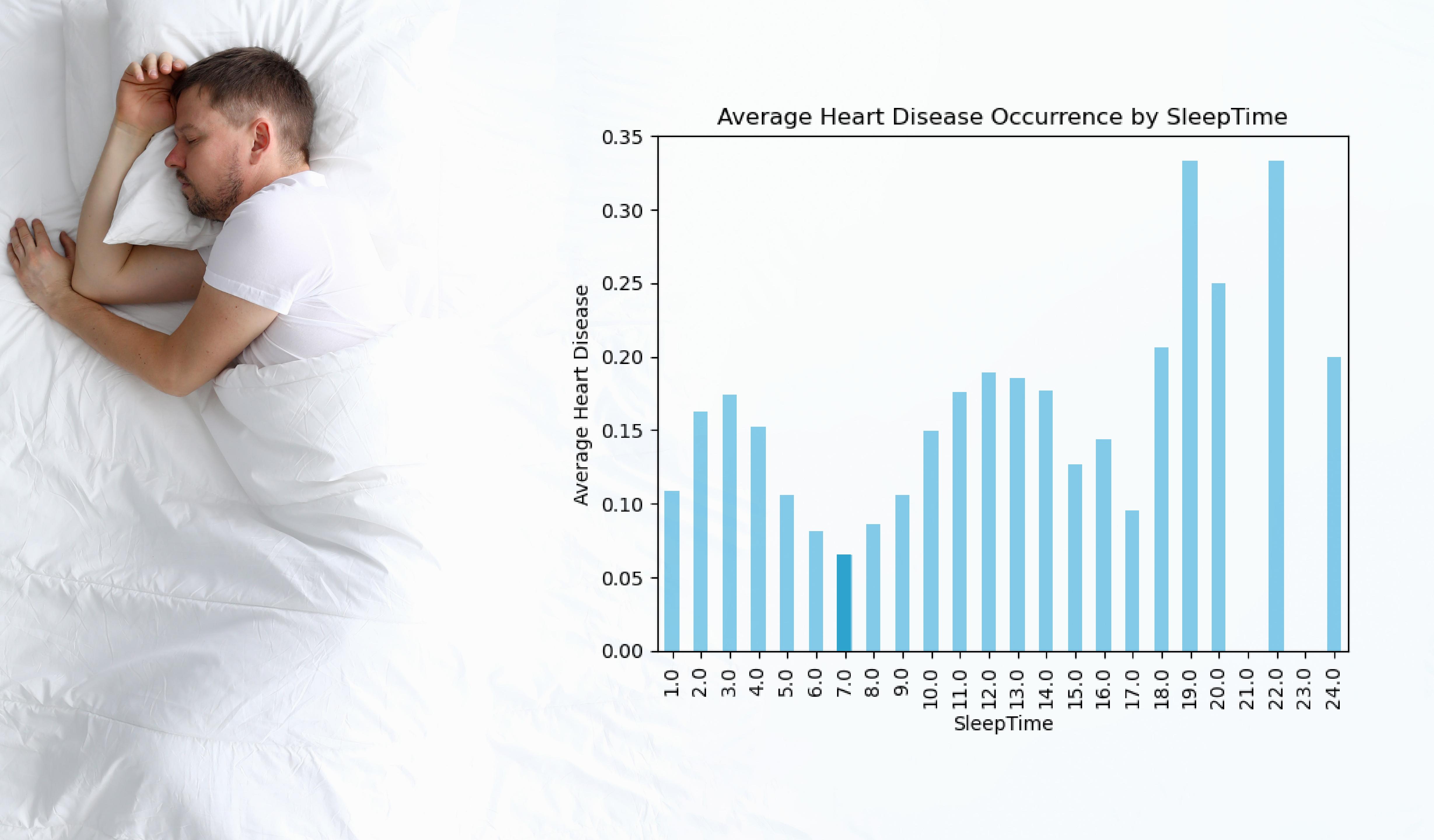


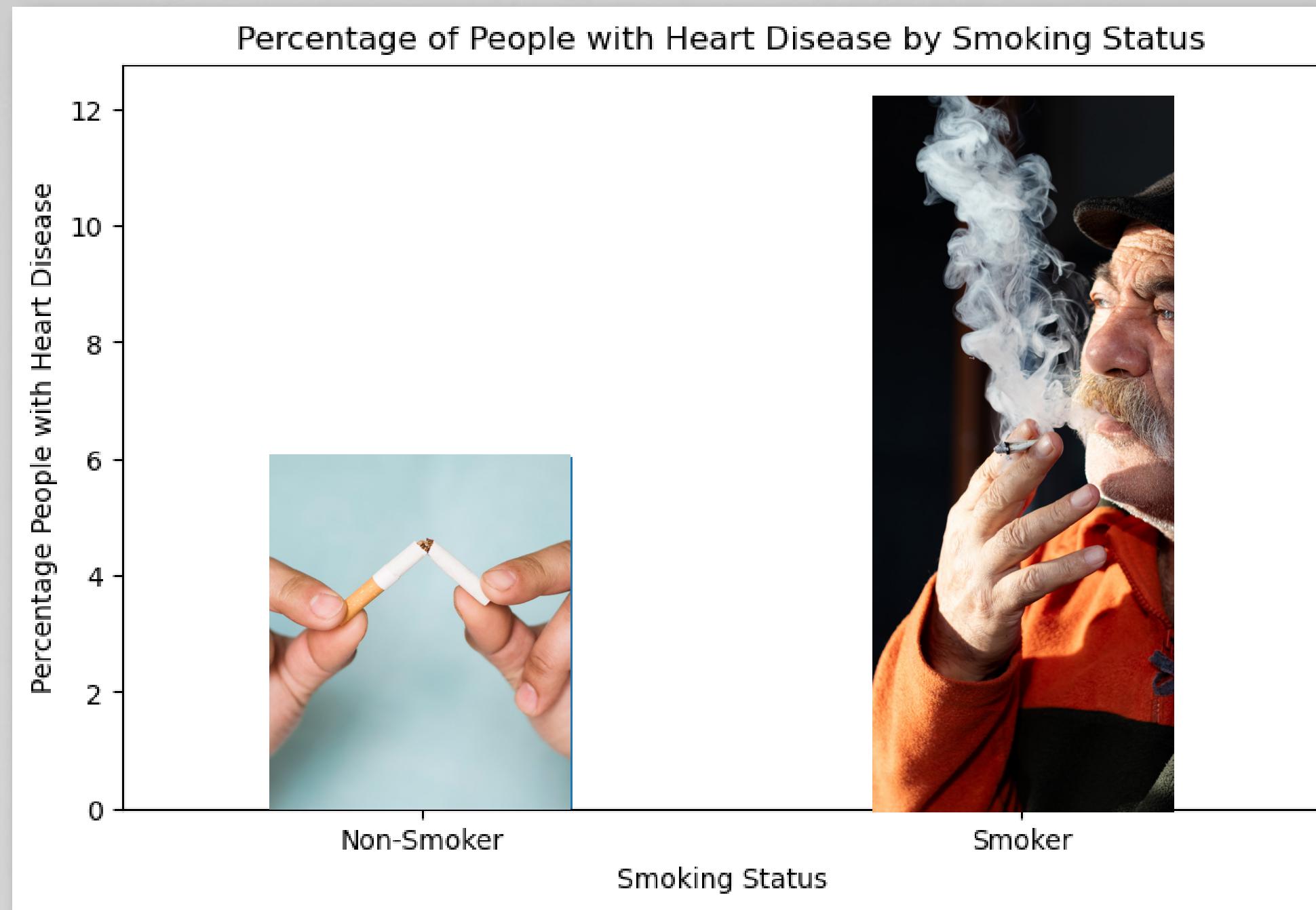
# Q4

**What could be the most important  
choices made to reverse the potential  
negative outcomes of heart disease?**

living an active and healthy lifestyle; perhaps that would include exercise, staying physically active, and maintaining your ability to walk as long as possible. Also, the data around BMI seems to suggest that staying around 20-25BMI has the least risk, so keeping a healthy diet would help in reducing your risk of heart disease



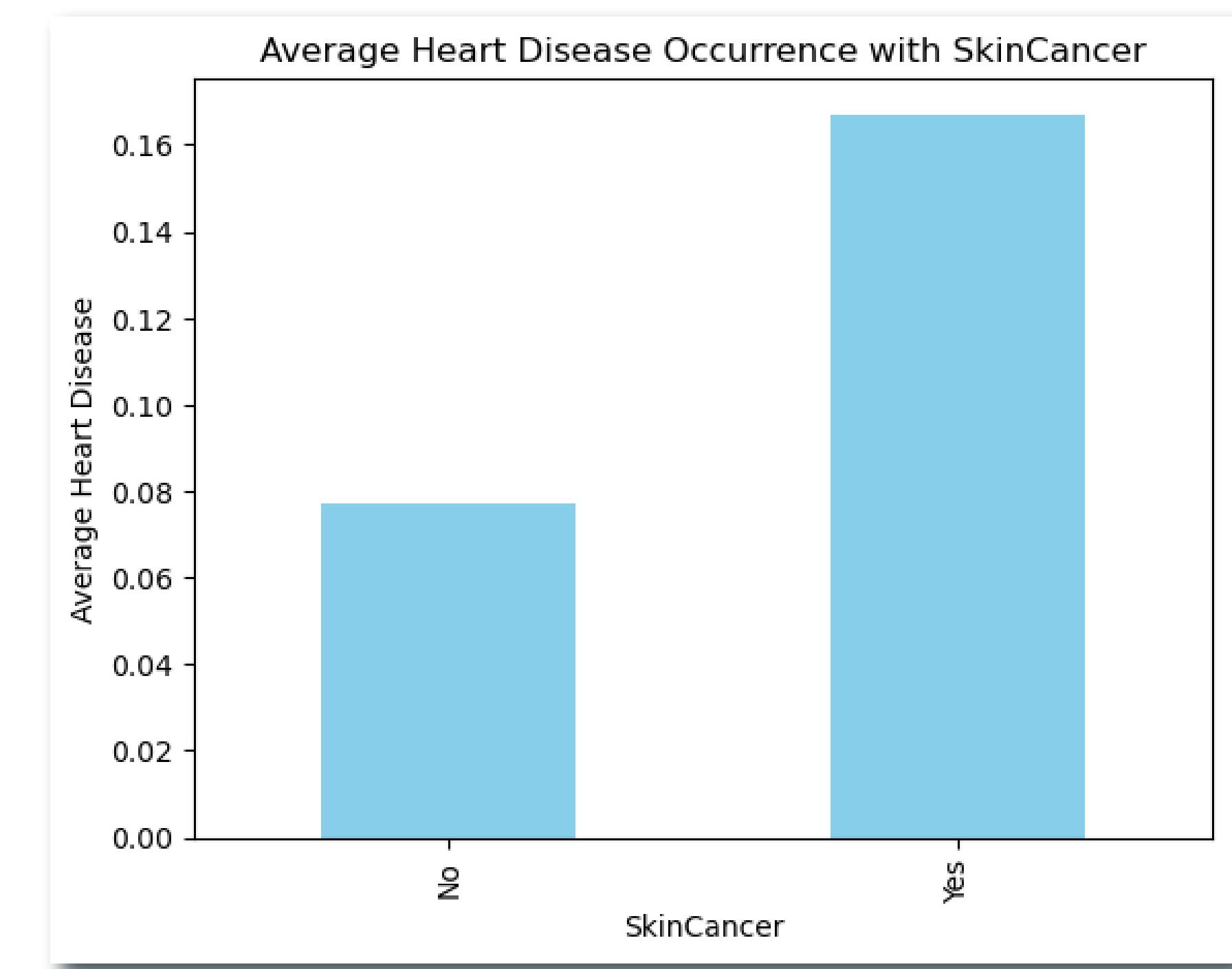


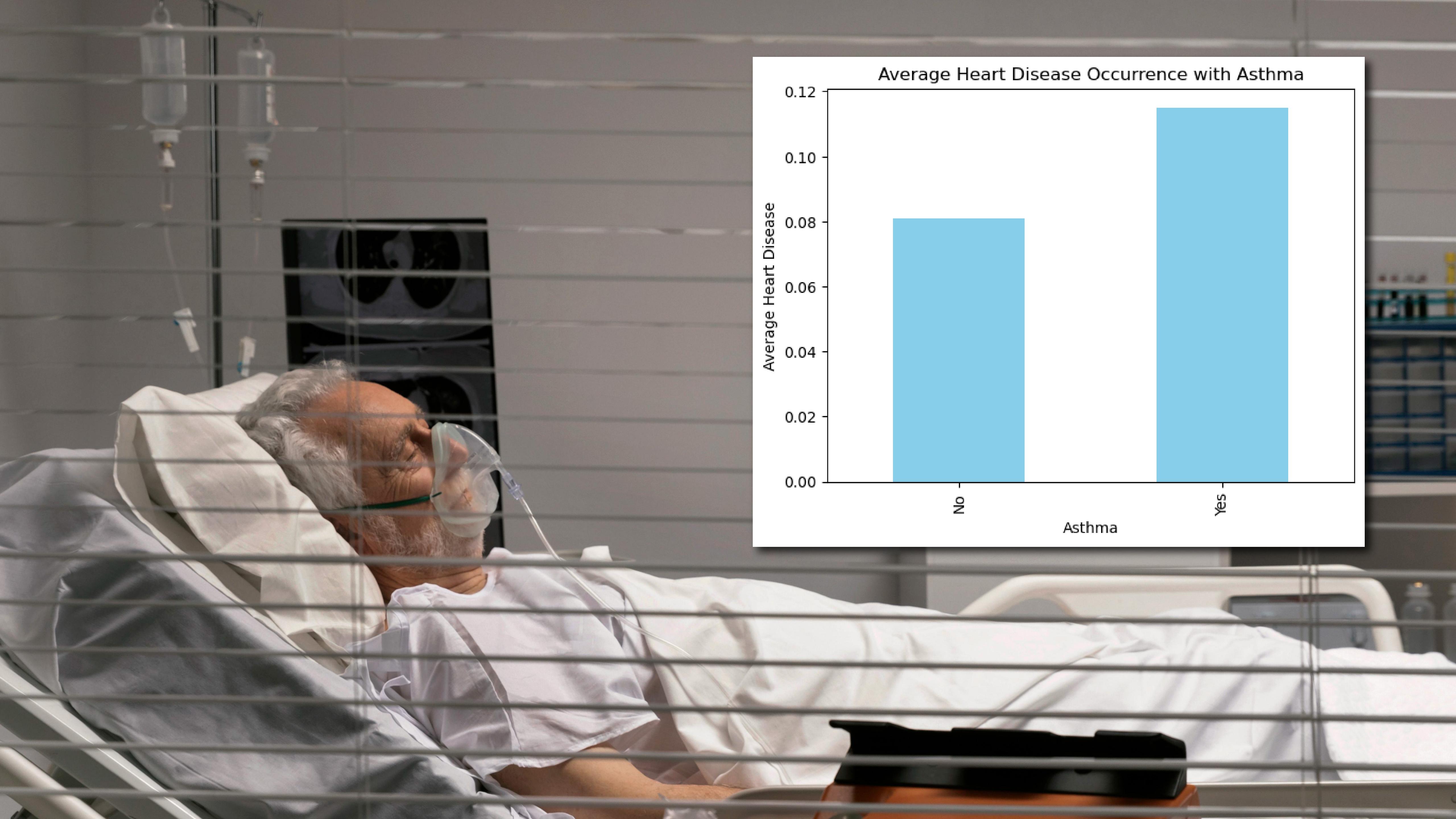


# Q5

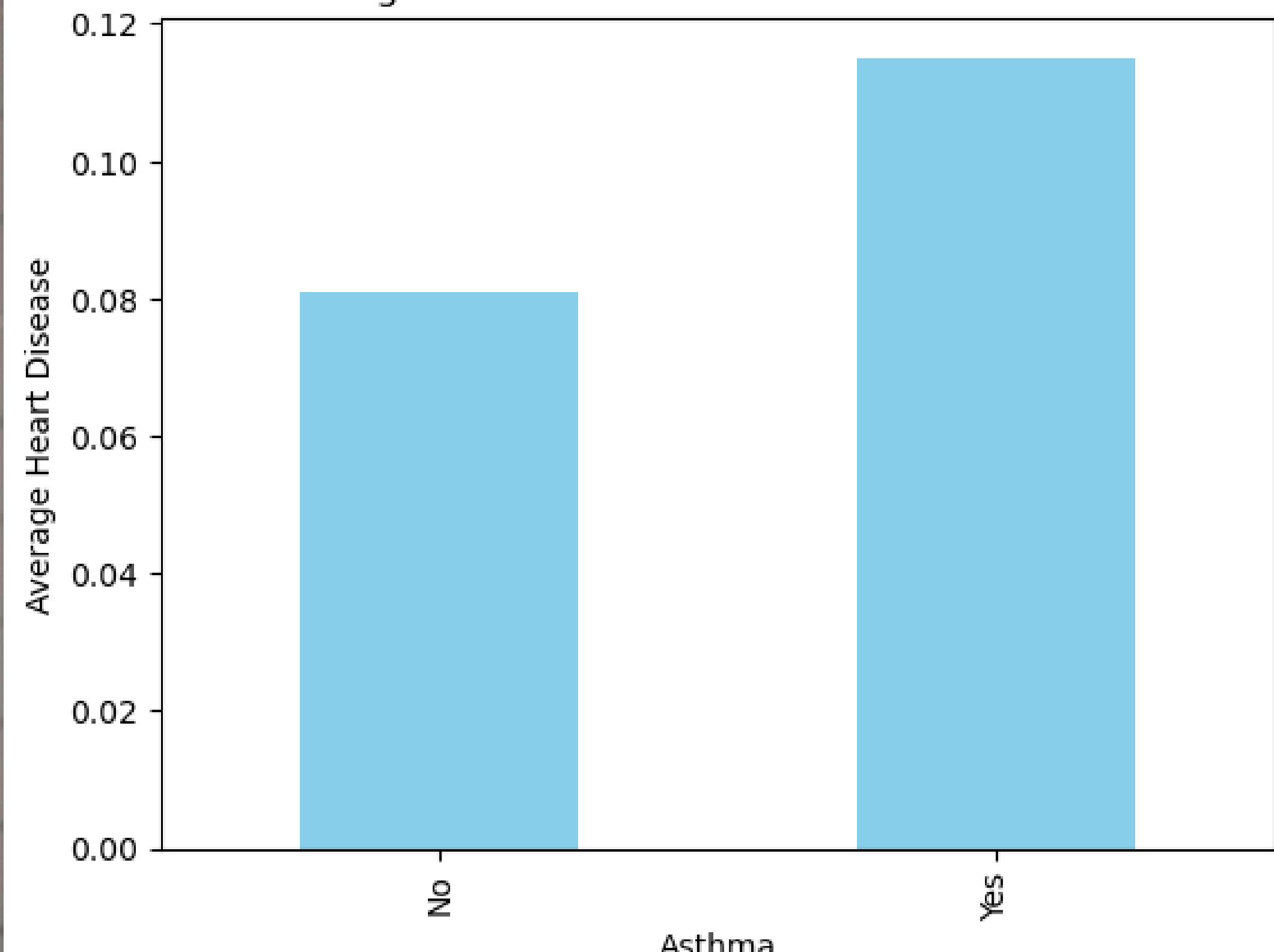
How will this data help hospitals to  
make **informed decisions** and develop  
effective business plans and strategies?

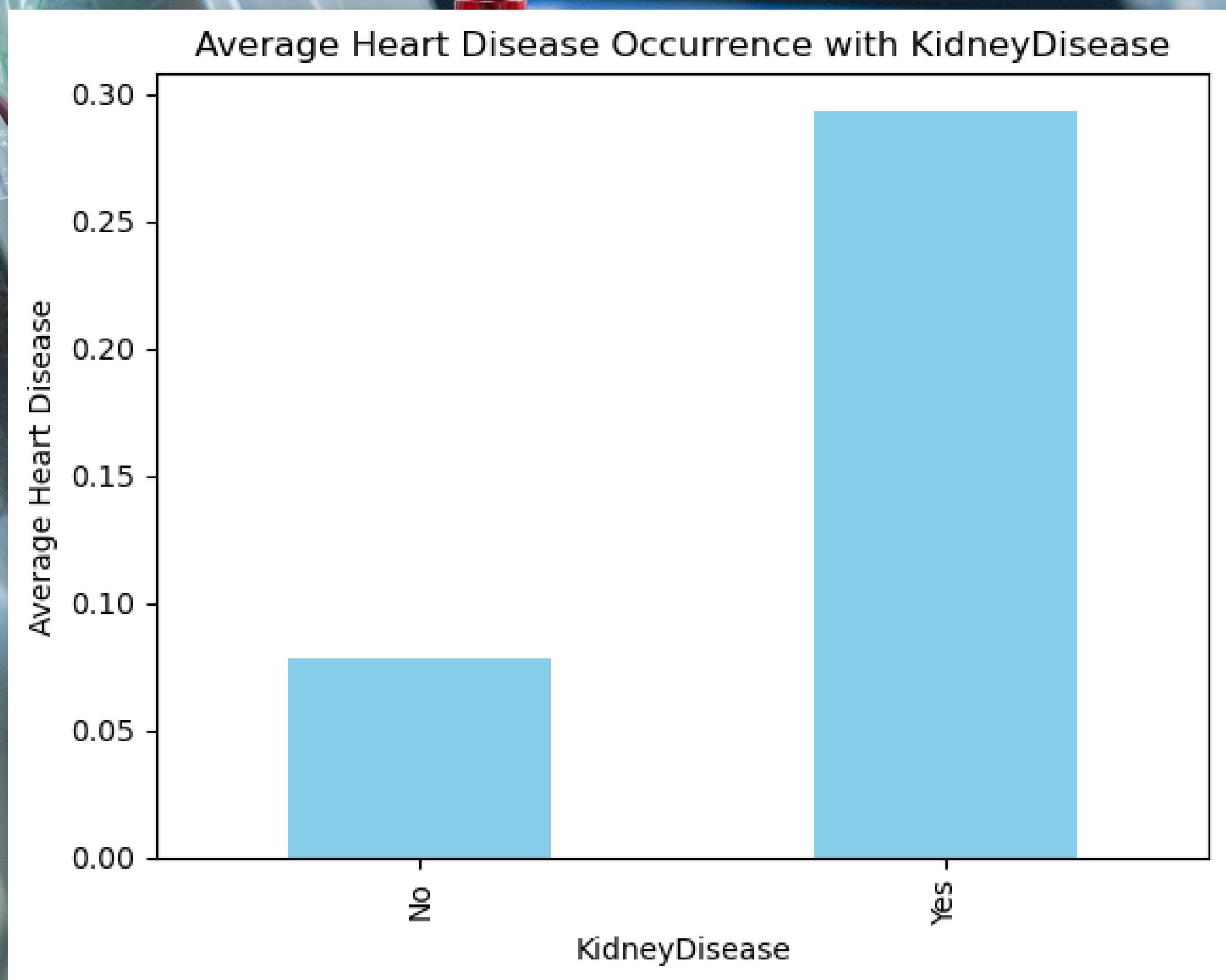
Some of the data surrounding the increases in heart disease as a comorbidity for other medical issues would also be important information for hospitals to have access to. For instance, people that have had skin cancers, asthma, diabetes, and kidney disease all had a higher risk of developing heart disease than those people who did not.





Average Heart Disease Occurrence with Asthma





# Credentials

- **Project Team:**

- Jesse Kranyak
- Jeff Boczkaja
- Mohamed Altoobi
- Seriesha Mandava

- **References:**

- all Photos/ images : <https://www.freepik.com>
- Color Palate : <https://color.adobe.com>
- Fonts used : <https://fonts.adobe.com>
- Software : <https://www.adobe.com>

# Thank You